Folding of the Protein Domain hbSBD

Maksim Kouza[1,2], Chi-Fon Chang[3], Shura Hayryan[2], Tsan-hung Yu[4], Mai Suan Li[1], Tai-huang Huang[4,5], and Chin-Kun Hu[2,6]

[1] *Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, 02-668 Warszawa, Poland*

[2] *Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan*

[3] *Genomics Research Center, Academia Sinica, Nankang, Taipei 11529, Taiwan*

[4] *Institute of Biomedical Sciences, Academia Sinica, Nankang, Taipei 11529, Taiwan*

[5] *Department of Physics, National Taiwan Normal University, Taipei 11718, Taiwan.*

[6] *National Center for Theoretical Sciences at Taipei, Physics Division, National Taiwan University, Taipei 10617, Taiwan*

ABSTRACT. The folding of the α-helice domain hbSBD of the mammalian mitochondrial branched-chain α-ketoacid dehydrogenase (BCKD) complex is studied by the circular dichroism technique in absence of urea. Thermal denaturation is used to evaluate various thermodynamic parameters defining the equilibrium unfolding, which is well described by the two-state model with the folding temperature $T_F = 317.8 \pm 1.95$ K and the enthalpy change $\Delta H_G = 19.67 \pm 2.67$ kcal/mol. The folding is also studied numerically using the off-lattice coarse-grained Go model and the Langevin dynamics. The obtained results, including the population of the native basin, the free energy landscape as a function of the number of native contacts and the folding kinetics, also suggest that the hbSBD domain is a two-state folder. These results are consistent with the biological function of hbSBD in BCKD.

Introduction

Understanding the dynamics and mechanism of protein folding remains one of the most challenging problems in molecular biology [1]. Single domain α proteins attract much attention of researchers because most of them fold faster than β and $\alpha\beta$ proteins [2, 3] due to relatively simple energy landscapes and one can, therefore, use them to probe main aspects of the funnel theory [4]. Recently, the study of this class of proteins becomes even more attractive because the one-state or downhill folding may occur in some small α proteins [5, 6, 7].

The mammalian mitochondrial branched-chain α-ketoacid dehydrogenase (BCKD) complex catalyzes the oxidative decarboxylation of branched-chain α-ketoacids derived from leucine, isoleucine and valine to give rise to branched-chain acyl-CoAs. In patients with inherited maple syrup urine disease, the activity of the BCKD complex is deficient, which is manifested by often fatal acidosis and mental retardation [8]. The BCKD multi-enzyme complex (4,000 KDa in size) is organized about a cubic 24-mer core of dihydrolipoyl transacylase (E2), with multiple copies of hetero-tetrameric decarboxylase (E1), a homodimeric dihydrogenase (E3), a kinase (BCK) and a phosphatase attached through ionic interactions. The E2 chain of the human BCKD complex, similar to other related multi-functional enzymes [9], consists of three domains: The amino-terminal lipoyl-bearing domain (hbLBD, 1-84), the interim E1/E3 subunit-binding domain (hbSBD, 104-152) and the carboxy-terminal inner-core domain. The structures of these domains serve as bases for modeling interactions of the E2 component with other components of α-ketoacid dehydrogenase complexes. The structure of hbSBD (Fig. 1) has been determined by NMR spectroscopy, and the main function of the hbSBD is to attach both E1 and E3 to the E2 core [10]. The two-helix structure of this domain is reminiscent of the small protein BBL [7] which may be a good candidate for observation of downhill folding [5, 6]. So the study of hbSBD is interesting not only because of the important biological role of the BCKD complex in human metabolism but also for illuminating folding mechanisms.

From the biological point of view, hbSBD could be less stable than hbLBD and one of our goals is, therefore, to check this by the circular dichroism (CD) experiments. In this paper we study the thermal folding-unfolding transition in the hbSBD by the CD technique in the absence of urea and pH=7.5. Our thermodynamic data do not show evidence for

the downhill folding and they are well fitted by the two-state model. We obtained folding temperature $T_F = 317.8 \pm 1.95$ K and the transition enthalpy $\Delta H_G = 19.67 \pm 2.67$ kcal/mol. Comparison of such thermodynamic parameters of hbSBD with those for hbLBD shows that hbSBD is indeed less stable as required by its biological function. However, the value of ΔH_G for hbSBD is still higher than those of two-state α proteins reported in [3], which indicates that the folding process in the hbSBD domain is highly cooperative.

From the theoretical point of view it is very interesting to establish if the two-state foldability of hbSBD can be captured by some model. The all-atom model would be the best choice for a detailed description of the system but the study of hbSBD requires very expensive CPU simulations. Therefore we employed the off-lattice coarse-grained Go-like model [11, 12] which is simple and allows for a thorough characterization of folding properties. In this model amino acids are represented by point particles or beads located at positions of C_α atoms. The Go model is defined through the experimentally determined native structure [10], and it captures essential aspects of the important role played by the native structure [12, 13].

It should be noted that the Go model by itself can not be employed to ascertain the two-state behavior of proteins. However, one can use it in conjunction with experiments providing the two-state folding because this model does not *always* provide the two-state behavior as have been clearly shown in the seminal work of Clementi *et al.* [12]. In fact, the Go model correctly captures not only the two-state folding of proteins CI2 and SH3 (more two-state Go folders may be found in Ref. [14]) but also intermediates of the three-state folder barnase, RNAse H and CheY [12]. The reason for this is that the simple Go model ignores the energetic frustration but it still takes the topological frustration into account. Therefore, it can capture intermediates that occur due to topological constraints but not those emerging from the frustration of the contact interactions.

With the help of Langevin dynamics simulations and the histogram method [15] we have shown that, in agreement with our CD data, hbSBD is a two-state folder with a well-defined transition state (TS) in the free energy landscape. The two helix regions were found to be highly structured in the TS. The two-state behavior of hbSBD is also supported by our kinetics study showing that the folding kinetics follows the single exponential scenario. The two-state folding obtained in our simulations suggests that for hbSBD the topological frustration is more important than the energetic factor.

The dimensionless quantity, Ω_c [16], which characterizes the structural cooperativity of

the thermal denaturation transition was computed and the reasonable agreement between the CD experiments and Go simulations was obtained. Incorporation of side chains may give a better agreement [16, 17] but this problem is beyond the scope of the present paper.

Materials and Methods

Sample Preparation

hbSBD protein was purified from the BL21(DE3) strain of *E. coli* containing a plasmid that carried the gene of hbLBD(1-84), a TEV cleavage site in the linker region, and hbSBD (104-152), generously provided to us by Dr. D.T. Chuang of the Southwestern Medical Center, University of Texas. There is an extra glycine in front of Glu104 which is left over after TEV cleavage, and extra leucine, glutamic acid at the C-terminus before six histidine residues. The protein was purified by Ni-NTA affinity chromatography, and the purity of the protein was found to be better than 95%, based on the Coomassie blue-stained gel. The complete sequence of $N = 52$ residues for hbSBD is
(G)EIKGRKTLATPAVRRLAMENNIKLSEVVGSGKDGRILKEDILNYLEKQT(L)(E).

Circular Dichroism

CD measurements were carried out in Aviv CD spectrometer model 202 with temperature and stir control units at different temperature taken from 260nm to 195nm. All experiments were carried at 1 nm bandwidth in 1.0 cm quartz square cuvette thermostated to $\pm 0.1^o$C. Protein concentration ($\sim$ 50 uM) was determined by UV absorbance at 280nm using ϵ_{280nm}=1280 $\mathrm{M}^{-1}\mathrm{cm}^{-1}$ with 50mM phosphate buffer at pH7.5. Temperature control was achieved using a circulating water bath system, and the equilibrium time was three minutes for each temperature point. The data was collected at each 2K increment in temperature. The study was performed at heating rate of 10^oC/min and equilibration time of 3 minutes. The volume changes as a result of thermal expansion as well as evaporation of water were neglected.

Fitting Procedure

Suppose the thermal denaturation is a two-state transition, we can write the ellipticity as

$$\theta \;=\; \theta_D + (\theta_N - \theta_D) f_N \,, \tag{1}$$

where θ_D and θ_N are values for the denaturated and folded states. The fraction of the folded

conformation f_N is expressed as [18]

$$\begin{aligned} f_N &= \frac{1}{1+\exp(-\Delta G_T/T)}, \\ \Delta G_T &= \Delta H_T - T\Delta S_T = \Delta H_G\left(1-\frac{T}{T_G}\right) \\ &+ \Delta C_p\left[(T-T_G) - T\ln\frac{T}{T_G}\right]. \end{aligned} \tag{2}$$

Here ΔH_G and ΔC_p are jumps of the enthalpy and heat capacity at the mid-point temperature T_G (also known as melting or folding temperature) of thermal transition, respectively. Some other thermodynamic characterization of stability such as the temperature of maximum stability (T_S), the temperature with zero enthalpy (T_H), and the conformational stability (ΔG_S) at T_S can be computed from results of regression analysis [19]

$$\ln\frac{T_G}{T_S} = \frac{\Delta H_G}{T_G\Delta C_p}, \tag{3}$$

$$T_H = T_G - \frac{\Delta H_G}{\Delta C_p}, \tag{4}$$

$$\Delta G_S = \Delta C_p(T_S - T_H). \tag{5}$$

Using Eq. (1) - Eq. (5) we can obtain all thermodynamic parameters from CD data.

It should be noted that the fitting of Eq. 2 with $\Delta C_p > 0$ allows for an additional cold denaturation [20] at temperatures much lower than the room temperature . The temperature of such a transition, T'_G, may be obtained by the same fitting procedure with an additional constraint of $\Delta H_G < 0$. Since the cold denaturation transition is not seen in Go models, to compare the simulation results to the experimental ones we also use the approximation in which $\Delta C_p = 0$.

Simulation

We use coarse-grained continuum representation for hbSBD protein, in which only the positions of 52 C_α-carbons are retained. We adopt the off-lattice version of the Go model [11] where the interaction between residues forming native contacts is assumed to be attractive and the non-native interactions - repulsive. Thus, by definition for the Go model the PDB structure is the native structure with the lowest energy. The advantage of this model is its simplicity which allows one to study model proteins in detail. Following Ref. 12, we write

the energy of the Go-like model as

$$
\begin{aligned}
E &= \sum_{bonds} K_r(r_{i,i+1} - r_{0i,i+1})^2 + \sum_{angles} K_\theta(\theta_i - \theta_{0i})^2 \\
&+ \sum_{dihedral} \{K_\phi^{(1)}[1 - \cos(\Delta\phi_i)] + K_\phi^{(3)}[1 - \cos 3(\Delta\phi_i)]\} \\
&+ \sum_{i>j-3}^{NC} \epsilon_1 \left[5R_{ij}^{12} - 6R_{ij}^{10}\right] + \sum_{i>j-3}^{NNC} \epsilon_2 \left(\frac{C}{r_{ij}}\right)^{12}. \qquad (6)
\end{aligned}
$$

Here $\Delta\phi_i = \phi_i - \phi_{0i}$, $R_{ij} = r_{0ij}/r_{ij}$; $r_{i,i+1}$, θ_i and ϕ_i stand for the ith bond length between the ith and $(i+1)$th residues, the bond angle between the $(i-1)$th and ith bonds and the dihedral angle around the ith bond, respectively. r_{ij} is the distance between the ith and jth residues. Subscript "0", "NC" and "NNC" refer to the native conformation, native contacts and non-native contacts, respectively. The first harmonic term keeps the chain connectivity, while the second and third terms represent the local angular interactions. Two last terms are non-local interactions, where the former includes native contact interactions and the latter is nonspecific repulsion between non-native pairs. We choose $K_r = 100\epsilon_H$, $K_\theta = 20\epsilon_H, K_\phi^{(1)} = \epsilon_H, K_\phi^{(3)} = 0.5\epsilon_H, \epsilon_1 = \epsilon_H, \epsilon_2 = \epsilon_H$ and $C = 4$ Å, where ϵ_H is the characteristic hydrogen bond energy [12].

The nativeness of any configuration is measured by the number of native contacts Q. We define that the ith and jth residues are in the native contact if r_{0ij} is smaller than a cutoff distance d_c taken to be $d_c = 7.5$ Å, where r_{0ij} is the distance between the ith and jth residues in the native conformation. Using this definition and the native conformation of Ref. 10, we found that the total number of native contacts Q_{total} is 62. To study the probability of being in the native state we use the following overlap function [21]

$$
\chi = \frac{1}{Q_{total}} \sum_{i<j+1}^{N} \theta(1.2r_{0ij} - r_{ij})\Delta_{ij} \qquad (7)
$$

where Δ_{ij} is equal to 1 if residues i and j form a native contact and 0 otherwise and $\theta(x)$ is the Heaviside function. The argument of this function guarantees that a native contact between i and j is classified as formed when r_{ij} is shorter than $1.2r_{0ij}$ [12]

The overlap function χ, which is one if the conformation of the polypeptide chain coincides with the native structure and zero for unfolded conformations, can serve as an order parameter for the folding-unfolding transition. The probability of being in the native state,

f_N, which can be measured by the CD and other experimental techniques, is defined as $f_N =< \chi >$, where $< ... >$ stands for a thermal average.

The dynamics of the system is obtained by integrating the following Langevin equation [22]

$$m\frac{d^2\vec{r}}{dt^2} = -\zeta\frac{d\vec{r}}{dt} + \vec{F}_c + \vec{\Gamma}, \quad (8)$$

where m is the mass of a bead, ζ is the friction coefficient, $\vec{F}_c = dE/d\vec{r}$. The random force $\vec{\Gamma}$ is a white noise, i.e. $< \Gamma_i(t)\Gamma_j(t') >= 2\zeta k_B T\delta_{ij}\delta(t-t')$, where i and j refer to components x, y and z. It should be noted that the folding thermodynamics does not depend on the enviroment viscosity (or on ζ) but the folding kinetics depends on it [23]. We chose the dimensionless parameter $\tilde{\zeta} = (\frac{a^2}{m\epsilon_H})^{1/2}\zeta = 8$, where m is the mass of a bead and a is the bond length between successive beads. One can show that this value of $\tilde{\zeta}$ belongs to the interval of the viscosity where the folding kinetics is fast. We have tried other values of $\tilde{\zeta}$ but the results remain unchanged qualitatively.

We measure time in units of $\tau_L = (ma^2/\epsilon_H)^{1/2}$. Using the typical value $m = 3.10^{-25}$ kg [24], $a = 4$Å and $\epsilon_H = 0.91$ kcal/mol (this choice of ϵ_H follows from the requirement that the simulated folding temperature coincides with the experimental one, see below) we obtained $\tau_L \approx 3$ ps. The dynamics Eq. (8) was solved by the Verlet algorithm [25] with the time step $\Delta t = 0.005\tau_L$. All thermodynamic quantities are obtained by the histogram method [15].

Results

CD Experiments

The structure of hbSBD is shown in Figure 1. Its conformational stability is investigated in present study by analyzing the unfolding transition induced by temperature as monitored by CD, similar to that described previously [26, 27]. The reversibility of thermal denaturation was ascertained by monitoring the return of the CD signal upon cooling from 95^{o}C to 22 ^{o}C; immediately after the conclusion of the thermal transition. The transition was found to be more than 80% reversible. Loss in reversibility to greater extent was observed on prolonged exposure of the sample to higher temperatures. This loss of reversibility is presumably due to irreversible aggregation or decomposition. Figure 2 shows the wavelength dependence of mean residue molar ellipticity of hbSBD at various temperatures between 278K and 363K.

In a separate study, the thermal unfolding transition as monitored by ellipticity at 228 nm was found to be independent of hbSBD concentration in the range of 2 uM to 10 uM. It was also found to be unaffected by change in heating rate between 2^oC/min to 20^oC/min. These observations suggest absence of stable intermediates in heat induced denaturation of hbSBD. A valley at around 220 nm, characteristics of the helical secondary structure is evident for hbSBD.

Figure 3 shows the temperature dependence of the population of the native conformation, f_N, for wave lengths $\lambda = 208, 212$ and 222 nm. We first try to fit these data to Eq. (2) with $\Delta C_p \neq 0$. The fitting procedure gives slightly different values for the folding (or melting) temperature and the enthalpy jump for three sets of parameters. Averaging over three values, we obtain $T_G = 317.8 \pm 1.95$ K and $\Delta H_G = 19.67 \pm 2.67$ kcal/mol. Other thermodynamic quantities are shown on the first row of Table 1. The similar fit but with $\Delta C_p = 0$ gives the thermodynamic parameters shown on the second row of this table. Since the experimental data are nicely fitted to the two-state model we expect that the downhill scenario does not applied to the hbSBD domain.

For the experimentally studied temperature interval two types of the two-state fit (2) with $\Delta C_p = 0$ and $\Delta C_p \neq 0$ give almost the same values for T_G, ΔH_G and ΔS_G. However, pronounced different behaviors of the population of the native basin, f_N, occur when we interpolate results to the low temperature region (Fig. 4). For the $\Delta C_p = 0$ case, f_N approaches the unity as $T \to 0$ but it goes down for $\Delta C_p \neq 0$. This means that the $\Delta C_p \neq 0$ fit is valid if the second cold denaturation transition may occur at T_G'. This phenomenon was observed in single domains as well as in multi-domain globular proteins [20]. We predict that the cold denaturation of hbSBD may take place at $T'_G \approx 212$ K which is lower than $T'_G \approx 249.8$ K for hbLBD shown on the 4th row of Table 1. It would be of great interest to carry out the cold denaturation experiments in cryo-solvent to elucidate this issue.

To compare the stability of the hbSBD domain with the hbLBD domain which has been studied in detail previously [27] we also present the thermodynamic data of the latter on Table 1. Clearly, hbSBD is less stable than hbLBD by its smaller ΔG_S and lower T_G values. This is consistent with their respective backbone dynamics as revealed by ^{15}N-T_1, ^{15}N-T_2, and ^{15}N-^{1}H NOE studies of these two domains using uniformly ^{15}N-labeled protein samples (Chang and Huang, unpublished results). Biologically, hbSBD must bind to either E1 or E3 at different stages of the catalytic cycle, thus it needs to be flexible to adapt to local

enviroments of the active sites of E1 and E3. On the other hand, the function of hbLBD is to permit its Lys44 residue to channel acetyl group between donor and acceptor molecules and only the Lys44 residue needs to be flexible [28]. In addition, the NMR observation for the longer fragment (comprising residues 1-168 of the E2 component) also showed that the hbLBD region would remain structured after several months while the hbSBD domain could de-grate in a shorter time.

Folding Thermodynamics from Simulations

In order to calculate the thermodynamics quantities we have collected histograms for the energy and native contacts at six values of temperature: $T = 0.4, 0.5, 0.6, 0.7, 0.8$ and 1.0 ϵ_H/k_B. For sampling, at each temperature 30 trajectories of 16×10^7 time steps have been generated with initial 4×10^7 steps discarded for thermalization. The reweighting histogram method [15] was used to obtain the thermodynamics parameters at all temperatures.

Figure 4 (open circles) shows the temperature dependence of population of the native state, defined as the renormalized number of native contacts (see Material and Methods) for the Go model. Since there is no cold denaturation for this model, to obtain the thermodynamic parameters we fit f_N to the two-state model (Eq. 2) with $\Delta C_p = 0$.

The fit (black curve) works pretty well around the transition temperature but it gets worse at high T due to slow decay of f_N which is characteristic for almost all of theoretical models. In fitting we have chosen the hydrogen bond energy $\epsilon_H = 0.91$ kcal/mol in Hamiltonian (6) so that $T_G = 0.7\epsilon_H/k_B$ coincides with the experimental value 317.8 K. From the fit we obtain $\Delta H_G = 11.46$ kcal/mol which is smaller than the experimental value indicating that the Go model is less stable compared to the real hbSBD.

Figure 5 shows the temperature dependence of derivative of the fraction of native contacts with respect to temperature df_N/dT (we also call this value the structural susceptibility) and the specific heat C_v obtained from the Go simulations. The collapse temperature T_θ, defined as the temperature at which C_v is maximal, almost coincides with the folding temperature T_F (at T_F the structural susceptibility has maximum). According to Klimov and Thirumalai [29], the dimensionless parameter $\sigma = \frac{|T_\theta - T_F|}{T_F}$ may serve as an indicator for foldablity of proteins. Namely, sequences with $\sigma \leq 0.1$ fold much faster that those which have the same number of residues but with σ exceeding 0.5. From this perspective, having $\sigma \approx 0$ hbSBD is supposed to be a good folder *in silico*. However, one has to be cautious about this conclusion

because the pronounced correlation between folding times τ_F and the equilibrium parameter σ, observed for simple on- and off-lattice models [24, 29] may be not valid for proteins in laboratory [30]. In our opinion, since the data collected from theoretical and experimental studies are limited, further studies are required to clarify the relationship between τ_F and σ.

Using experimental values for T_G (as T_F) and ΔH_G and the two-state model with $\Delta C_p = 0$ (see Table 1) we can obtain the temperature dependence of the population of native state f_N and, therefore, df_N/dT for hbSBD (Fig. 5). Clearly, the folding-unfolding transition *in vitro* is sharper than in the Go modeling. One of possible reasons is that our Go model ignores the side chain which can enhance the cooperativity of the denaturation transition [16].

The sharpness of the fold-unfolded transition might be characterized quantitatively via the cooperativity index Ω_c which is defined as follows [31]

$$\Omega_c = \frac{T_F^2}{\Delta T}\left(\frac{df_N}{dT}\right)_{T=T_F}, \tag{9}$$

where ΔT is the transition width. From Fig. 5, we obtain $\Omega_c = 51.6$ and 71.3 for the Go model and CD experiments, respectively. Given the simplicity of the Go model used here the agreement in Ω_c should be considered reasonable. We can also estimate Ω_c from the scaling law suggested in Ref. 32, $\Omega_c = 0.0057 \times N^{\mu}$, where exponent μ is universal and expressed via the random walk susceptibility exponent γ as $\mu = 1 + \gamma \approx 2.22 (\gamma \approx 1.22)$. Then we get $\Omega_c \approx 36.7$ which is lower than the experimental as well as simulation result. This means that hbSBD *in vitro* is, on average, more cooperative than other two-state folders.

Another measure for the cooperativity is κ_2 which is defined as [33] $\kappa_2 = \Delta H_{vh}/\Delta H_{cal}$, where $\Delta H_{vh} = 2T_{max}\sqrt{k_B C_V(T_{max})}$ and $\Delta H_{cal} = \int_0^\infty C_V(T)dT$, are the van't Hoff and the calorimetric enthalpy, respectively, $C_V(T)$ is the specific heat. Without the baseline substraction in $C_V(T)$ [34], for the Go model of hbSBD we obtained $\kappa_2 \approx 0.25$. Applying the baseline substraction as shown in the lower part of Fig. 5 we got $\kappa_2 \approx 0.5$ which is still much lower than $\kappa_2 \approx 1$ for a trully all-or-none transition. Since κ_2 is an extensive parameter, its low value is due to the shortcomings of the off-lattice Go models but not due to the finite size effects. More rigid lattice models give better results for the calorimetric cooperativity [17]. Thus, for the hbSBD domain the Go model gives the better agreement with our CD experiments for the structural cooperativity Ω_c than for the calorimetric measure κ_2.

Free Energy Profile

To get more evidence that hbSBD is a two-state folder we study the free energy profile using some quantity as a reaction coordinate. The precise reaction coordinate for a multi-dimensional process such as protein folding is difficult to ascertain. However, Onuchic and coworkers [35] have argued that, for minimally frustrated systems such as Go models, the number of native contact Q may be appropriate. Fig. 6(a) shows the dependence of free energy on Q for $T = T_F$. Since there is only one local maximum corresponding to the transition state (TS), hbSBD is a two-state folder. This is not unexpectable for hbSBD which contains only helices. The fact that the simple Go model correctly captures the two-state behavior as was observed in the CD experiments, suggests that the energetic frustration ignored in this model plays a minor role compared to the topological frustration [12].

We have sorted out structures of the denaturated state (DS), TS and the folded state (FS) at $T = T_F$ generating 10^4 conformations in equilibrium. The distributions of the RMSD, $P_{\rm RMSD}$, of these states are plotted in Fig. 6(b). As expected, $P_{\rm RMSD}$ for the DS spreads out more than that for the TS and FS. According to the free energy profile in Fig. 6(a), the TS conformations have 26 - 40 native contacts. We have found that the size (number of folded residues) [36] of the TS is equal to 32. Comparing this size with the total number of residues ($N = 52$) we see that the fraction of folded residues in the TS is higher than the typical value for real two-state proteins [36]. This is probably an artefact of Go models [37]. The TS conformations are relatively compact having the ratio $< R_g^{TS} > / R_g^{NS} \approx 1.14$, where $< R_g^{TS} >$ is the average radius of gyration of the TS ensemble and R_g^{NS} is the radius of gyration of the native conformation shown in Fig. 1. Since the RMSD, calculated only for two helices, is about 0.8 Åthe structures of two helices in the TS are not distorted much. It is also evident from the typical structure of the TS shown in Fig.6(b) where the helix regions H_1 and H_2 involve residues 13 - 19 and 39 - 48, respectively (a residue is considered to be in the helix state if its dihedral angle is about 60^o). Note that H_1 has two residues less compared to H_1 in the native conformation (see the caption to Fig. 1) but H_2 has even one bead more than its native state counterpart. Overall, the averaged RMSD of the TS conformations from the native conformation (Fig. 1) is about 4.9 Å indicating that the TS is not close to the native one. As seen from Figs. 6(a) and 1, the main difference comes from the tail parts. The most probable conformations (corresponding to maximum of $P_{\rm RMSD}$ in

Fig. 6(b) of the FS have RMSD about 2.5 Å. This value is reasonable from the point of view of the experimental structure resolution.

Folding Kinetics

The two-state foldability, obtained from the thermodynamics simulations may be also probed by studying the folding kinetics. For this purpose we monitored the time dependence of the fraction of unfolded trajectories $P_u(t)$ defined as follows [38]

$$P_u(t) \; = \; 1 - \int_0^t P_{fp}^N(s)ds, \tag{10}$$

where P_{fp}^N is the distribution of first passage folding times

$$P_{fp}^N \; = \; \frac{1}{M}\sum_{i=1}^{M}\delta(s-\tau_{f,1i}). \tag{11}$$

Here $\tau_{f,1i}$ is time for the ith trajectory to reach the native state for the first time, M is the total number of trajectories used in simulations. A trajectory is said to be folded if all of native contacts form. As seen from Eqs 10 and 11, $P_u(t)$ is the fraction of trajectories which do not reach the native state at time t. In the two-state scenario the folding becomes triggered after overcoming only one free energy barrier between the transition state and the denaturated one. Therefore, $P_u(t)$ should be a single exponential, i.e. $P_u(t) \sim \exp(-t/\tau_F)$ (a multi-exponential behavior occurs in the case when the folding proceeds via intermediates) [38]. Since the function $P_u(t)$ can be measured directly by a number of experimental techniques [39, 40], the single exponential kinetics of two-state folders is supported by a large body of experimental work (see, i.e. Ref. [26] and references there).

Fig. 7 shows the semi-logarithmic plot for $P_u(t)$ at $T = T_F$ for the Go model. Since the single exponential fit works pretty well, one can expect that intermediates do not occur on the folding pathways. Thus, together with the thermodynamics data our kinetic study supports the two-state behavior of the hbSBD domain as observed on the CD experiments.

From the linear fit in Fig. 7 we obtain the folding time $\tau_F \approx 0.1\mu$s. This value is consistent with the estimate of the folding time defined as the average value of the first passage times. If we use the empirical formula for the folding time $\tau_F = \tau_F^0 \exp(1.1N^{1/2})$, where prefactor $\tau_F^0 = 0.4\mu$s and N is a number of amino acids [31] then $\tau_F = 1.1 \times 10^3\mu$s for $N = 52$. This value is about four orders of magnitude larger than that obtained from

the Go model. Thus the Go model can capture the two-state feature of the denaturation transition for hbSBD domain but not folding times.

Discusion

We have used CD technique and the Langevin dynamics to study the mechanism of folding of hbSBD. Our results suggest that this domain is a two-state folder. The CD experiments reveal that the hbSBD domain is less stable than the hbLBD domain in the same BCKD complex, but it is more stable and cooperative compared to other fast folding α proteins.

Both the thermodynamics and kinetics results, obtained from the Langevin dynamics simulations, show that the simple Go model correctly captures the two-state feature of folding. It should be noted that the two-state behavior is not the natural consequence of the Go modeling because it allows for fishing folding intermediates caused by the topological frustration. From this standpoint it may be used to decipher the foldability of model proteins for which the topological frustration dominates. The reasonable agreement between the results obtained by the Go modeling and our CD experiments, suggests that the native state topology of hbSBD is more important than the energetic factor.

The theoretical model gives the reasonable agreement with the CD experimental data for the structural cooperativity Ω_c. However, the calorimetric cooperativity criterion $\kappa_2 \approx 1$ for two-state folders is hard to fulfill within the Go model. From the $\Delta C_p \neq 0$ fitting procedure we predict that the cold denaturation of hbSBD may occur at $T \approx 212$ K and it would be very interesting to verify this prediction experimentally. We are using the package SMMP [41, 42] and a parallel algorithm [43] to perform all-atom simulation of hbSBD to check the relevant results.

This work was supported by KBN grant number 1P03B01827 and by National Science Council in Taiwan under grant numbers No. NSC 93-2112-M-001-027 (to CKH) and No. NSC 92-2113-M-001-056 (to THH). The NMR spectra used to determine hbSBD structure were obtained at the High-field NMR Core Facility, National Research Program for Genomic Medicine (NRPGM), Taiwan, Republic of China.

[1] Daggett, V. & Fersht A. R. (2003) Is there a unifying mechanism for protein folding? *Trends in Biochem. Sci.* **28**, 18-25.

[2] Jackson, S. E. (1998) How do small single domain proteins fold? *Des* **3**, R81-R91.

[3] Kubelka, J., Hofrichter, J. & Eaton, W. A. (2004) The protein folding 'speed limit', *Curr Opin Struct Biol* **14**, 76-88.

[4] Bryngelson,J., Onuchic,J. N., Socci, N. D. & Wolynes, P. G. (1995) Funnels, pathways, and the energy landscape of protein folding - synthesis. *Proteins: Struct. Funct. Genet.* **21**, 167-195.

[5] Garcia-Mira, M. M., Sadqi, M., Fischer, N. Sanchez-Ruiz, J. M. & Munoz, V. (2002) Experimental identification of downhill protein folding. *Science* **298**, 2191-2195.

[6] Oliva, F. Y. & Munoz, V. (2004) A simple thermodynamic test to discriminate between two-state and downhill folding. *J. Am. Chem. Soc.* **126**, 8596-8597.

[7] Ferguson,N., Schartau, P. J., Sharpe, T. D., Sato, S. & Fersht, A. R. (2004) One-state downhill versus conventional protein folding. *J. Mol. Biol.* **344**, 295.

[8] Chuang, D.T. & Shih, V. E. (2001) in The Metabolic and Molecular Basis of inherited Disease, eds. Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D. (McGraw-Hill, New York), pp.1971-2006.

[9] Perham, R.N (2000) Swinging Arms and Swinging Domains in Multifunctional Enzymes: Catalytic Machines for Multistep Reactions. *Annu. Rev. Biochem.* **69**, 961-1004.

[10] C.-F. Chang, Y.-J. Lin, T.A. Yu, J.L. Chuang, D. T.Chuang , and T.-h. Huang (2005) in preparation.

[11] Go, N. (1983) Theoretical studies of protein folding. *Ann. Rev. Biophys. Bioeng.* **12** No. 1, 183-210.

[12] Clementi, C., Nymeyer H. & Onuchic, J. (2000) Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **298**, 937-953.

[13] Takada, S. (1999) *Go*-ing for the prediction of protein folding mechanisms. *Proc. Natl. Acad. Sci. USA* **96**, 11698-11700.

[14] Koga, N. & Takaga, S. (2001) Roles of native topology and chain-length scaling in protein folding: A simulation study with a Go-like model. *J. Mol. Biol.* **313**, 171-180.

[15] Ferrenberg, A. M. & Swendsen, R. H. (1989) Optimized Monte-Carlo data analysis. *Phys. Rev. Lett.* **63**, 1195-1198.

[16] Klimov, D. K. & Thirumalai, D. (1998) Cooperativity in protein folding: from lattice models with sidechains to real proteins. *Fold. Des.* **3**, 127-139.

[17] Li, M. S., Klimov, D. K. & Thirumalai, D. (2005) Finite size effects on calorimetric cooperativity of two-state proteins. *Physica A* **350**, 38-44.

[18] Privalov, P. L. (1979) Stability of proteins: Small globular proteins. *Adv. Prot. Chem.* **33**, 167-241.

[19] Becktel, W. J. & Schellman, J. A. (1987) Protein stability curves. *Biopolymers* **26**, 1859-1877.

[20] Privalov, P. L. (1990) Cold denaturation of proteins. *Crit. Rev. Biochem. Mol. Biol.* **25**, 281-305.

[21] Camacho, C. J. & Thirumalai, D. (1993) Kinetics and thermodynamics of folding in model proteins. *Proc. Natl. Acad. Sci. USA* **90** 6369-6372.

[22] Allen, M. P. & Tildesley, D. J. (1987) *Computer simulations of liquids*, (Oxford Science Pub., Oxford, UK).

[23] Klimov, D. K. & Thirumalai, D. (1997) Viscosity dependence of the folding rates of proteins. *Phys. Rev. Lett.* **79**, 317-320.

[24] Veitshans, T., Klimov, D. K. & Thirumalai, D. (1997) Protein folding kinetics: time scales, pathways and energy landscapes in terms of sequence dependent properties. *Folding and Design* **2**, 1-22.

[25] Swope, W. C., Andersen, H. C., Berens, P. H. & Wilson, K. R. (1982) Computer simulation method for the calculation of equilibrium constants for the formation of physical clusters and molecules: Application to small water clusters. *J. Chem. Phys.* **76**, 637-649.

[26] Naik,M., Chang, Y.-C. & Huang, T.-h. Folding kinetics of the lipoic acid-bearing domain of human mitochondrial branched chain alpha-ketoacid dehydrogenase complex. (2002) *FEBS Lett.* **530**, 133-138.

[27] Naik, M. & Huang, T.-h. (2004) Conformational stability and thermodynamic characterization of the lipoic acid bearing domain of human mitochondrial branched chain alpha-ketoacid dehydrogenase. *Protein Sci.* **13**, 2483-2492.

[28] Chang, C.-F., Chou, H.-T., Chuang,J. L., Chuang, D. T. & Huang, T.-h. (2002) Solution structure and dynamics of the lipoic acid-bearing domain of human mitochondrial branched-

chain alpha-keto acid dehydrogenase complex. *J. Bio. Chem* **277**, 15865-15873.

[29] Klimov, D. K. & Thirumalai, D. (1996) Criterion that determines the foldability of proteins. *Phys. Rev. Lett.* **76**, 4070-4073.

[30] Gillepse, B. & Plaxco, K. W. (2004) Using protein folding rates to test protein folding theories. *Annu. Rev. Biochem.* **73**, 837-859.

[31] Li, M. S., Klimov, D.K. & Thirumalai,D. (2004) Thermal denaturation and folding rates of single domain proteins: size matters. *Polymer* **45**, 573-579.

[32] Li, M. S., Klimov, D. K. & Thirumalai, D. (2004) Finite size effects on thermal denaturation of globular proteins. *Phys. Rev. Lett.* **93**, 268107.

[33] Kaya H & Chan, H. S. (2000) Energetic components of cooperative protein folding. *Phys. Rev. Lett.* **85**, 4823-4826.

[34] Chan, H. S., Shimizu, S. & Kaya, H (2004) Cooperativity principles in protein folding. *Methods in Enzymology* **380**, 350-379.

[35] Nymeyer, H., Garcia, A. E. & Onuchic,J.N. (1998) Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc. Natl. Acad. Sci. USA* **95** , 5921-5926.

[36] Bai, Y., Zhou, Y., Zhou, H. (2004) Critical nucleation size in the folding of small apparently two-state proteins. *Protein Sci.* **13**, 1173-1181.

[37] Li, M. S. & Kouza, M. (2005) in preparation.

[38] Klimov, D. K. & Thirumalai, D. (1999) Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. *Curr. Opin. Struct. Biol.* **97**, 97-107.

[39] Greene, L. H. ed. (2004). Elsevier Science. Investigating Protein Folding, Misfolding and Non-native States: Experimental and Theoretical Methods. *Methods* **34**

[40] Dyson, H. J. & Wright, P. E. (2005) Elucidation of the protein folding landscape by NMR. *Methods Enzymol.* **394**, 299-321.

[41] Eisenmenger,F., Hansmann, U.H.E., Hayryan, S. & Hu,C.-K. (2001) [SMMP] A modern package for simulation of proteins. *Comput. Phys. Commun.* **138**, 192-212.

[42] Eisenmenger,F., Hansmann, U.H.E., Hayryan, S. & Hu,C.-K. (2005) in preparation.

[43] Hayryan,S., Hu,C.-K., Hu, S.-Y. & and R.-J. Shang, R. J. (2001) Multicanonical parallel simulations of proteins with continuous potentials. *J. Comput. Chem.* **22**, 1287-1296.

Domain	$T_G(K)$	ΔH_G (kcal/mol/K)	ΔC_p (kcal/mol/K)	ΔS_G (cal/mol/K)	$T_S(K)$	$T_H(K)$	ΔG_S (kcal/mol)	$T'_G(K)$
SBD(exp)	317.8 ± 1.9	19.67 ± 2.67	0.387 ± 0.054	61.64 ± 7.36	270.9 ± 2.0	267.0 ± 2.1	1.4 ± 0.1	212 ± 2.5
SBD(exp)	317.9 ± 2.2	20.02 ± 3.11	0.0	62.96 ± 9.92	–	–	–	-
SBD(sim)	317.9 ± 7.95	11.46 ± 0.29	0.0	36.05 ± 1.85	–	–	–	-
LBD(exp)	344.0 ± 0.2	78.96 ± 1.28	1.51 ± 0.04	229.5 ± 3.7	295.7 ± 3.7	291.9 ± 1.3	5.7 ± 0.2	249.8 ± 1.1

TABLE 1. Thermodynamic parameters obtained from the CD experiments and simulations for hbSBD domain. The results shown on the first and fourth rows were obtained by fitting experimental data to the two-state equation (2) with $\Delta C_p \neq 0$. The second and third rows corresponding to the fit with $\Delta C_p = 0$. The results for hbLBD are taken from Ref. 27 for comparison.

Figure Captions

FIGURE 1. Ribbon representation of the structure of hbSBD domain. The helix region H_1 and H_2 include residues Pro12 - Glu20 and Lys39 - Glu47, respectively.

FIGURE 2. Dependence of the mean residue molar ellipticity on the wave length for 18 values of temperatures between 278 and 363 K.

FIGURE 3. Temperature dependence of the fraction of folded conformations f_N, obtained from the ellipticity θ by Eq. (2), for wave lengths $\lambda = 208$ (blue circles), 212 (red squares) and 222 nm (green diamonds). The solid lines corresponds to the two state fit given by Eq. 2 with $\Delta C_p \neq 0$. We obtained $T_G = T_F = 317.8 \pm 1.9$ K, $\Delta H_G = 19.67 \pm 2.67$ kcal/mol and $\Delta C_p = 0.387 \pm 0.054$.

FIGURE 4. The dependence of f_N for various sets of parameters. The blue and red curves correspond to the thermodynamic parameters presented on the first and the second rows of Table 1, respectively. Open circles refer to simulation results for the Go model. The solid black curve is the two-state fit ($\Delta C_p = 0$) which gives $\Delta H_G = 11.46$ kcal/mol and $T_F = 317.9$.

FIGURE 5. The upper part refers to the temperature dependence of df_N/dT obtained by the simulations (red) and the CD experiments (blue). The experimental curve is plotted using two-state parameters with $\Delta C_p = 0$ (see, the second row on Table 1). The temperature dependence of the heat capacity $C_V(T)$ is presented in the lower part. The dotted lines illustrate the base line substraction. The results are averaged over 20 samples.

FIGURE 6. (a) The dependence of free energy on the number of native contacts Q at $T = T_F$. The typical structures of the denaturated state , transition state and folded state are also drawn. The helix regions H_1 (green) and H_2 (orange) of the TS structure involve residues 13 - 19 and 39 - 48, respectively. For the FS structure H_2 is the same as for the TS structure but H_1 has two residues more (13 - 21). (b) Distributions of RMSD for three ensembles shown in (a). The average values of RMSD are equal to 9.8, 4.9 and 3.2 Å for the DS, TS and FS, respectively.

FIGURE 7. The semi-logarithmic plot of the time dependence of the fraction of unfolded trajectories at $T = T_F$. The distribution $P_u(t)$ was obtained from first passage times of 400 trajectories, which start from random conformations. The straight line corresponds to the fit $\ln P_{\rm u}(t) = -t/\tau_F$, where $\tau_F = 0.1\mu$s.

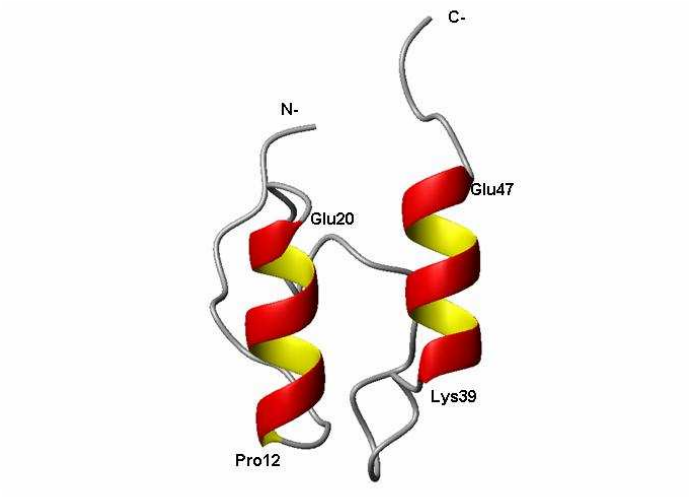

FIGURE 1.

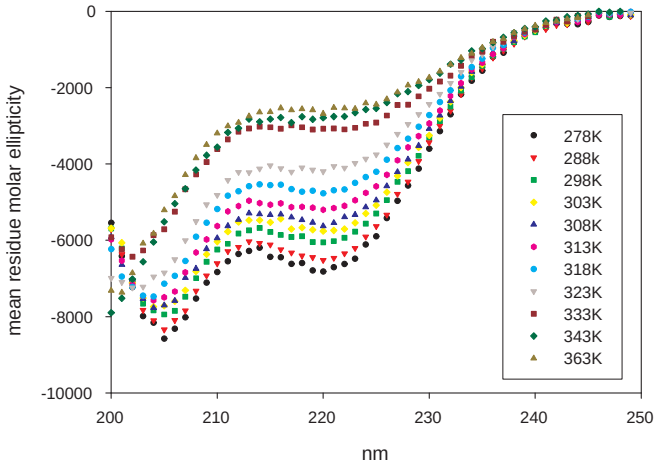

FIGURE 2.

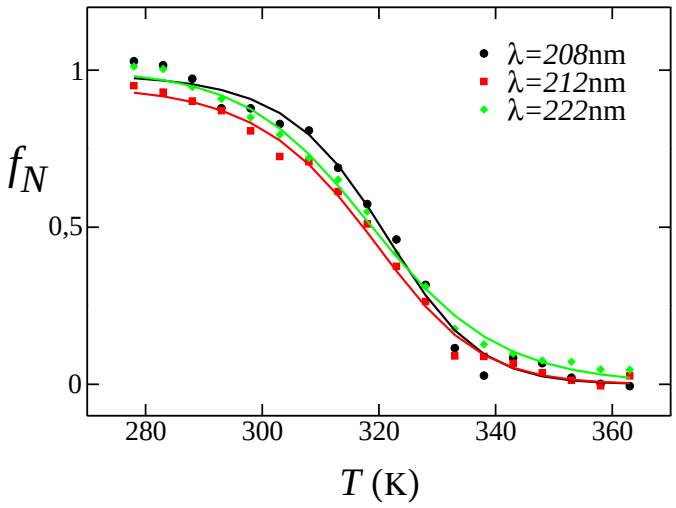

FIGURE 3.

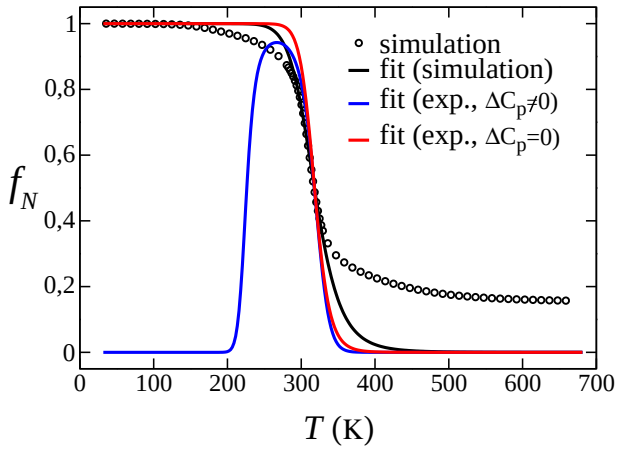
simulation
fit (simulation)
fit (exp., ΔC_p≠0)
fit (exp., ΔC_p=0)
f_N
T (K)
0
0,2
0,4
0,6
0,8
1
0
100
200
300
400
500
600
700

FIGURE 4.

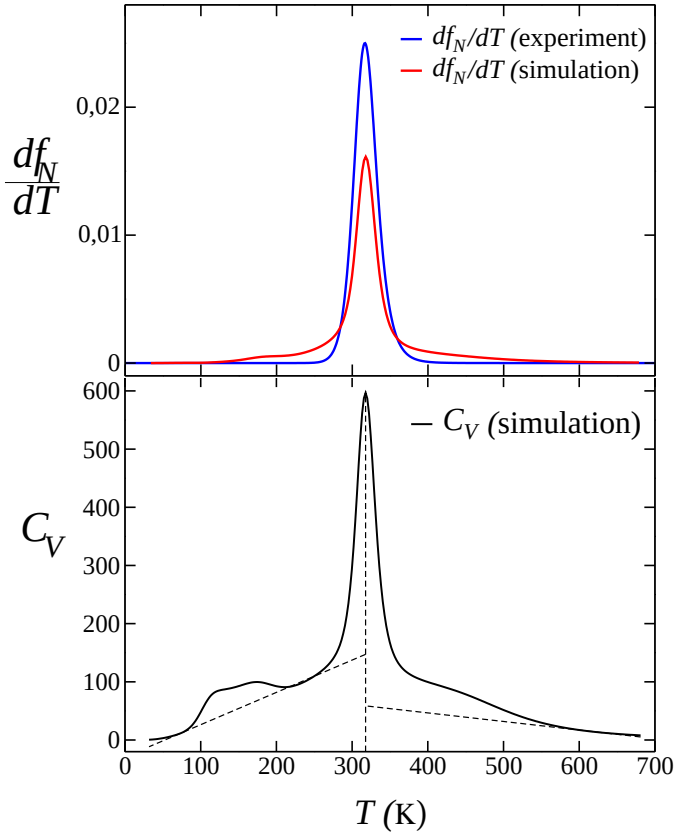
df_N/dT (experiment)
df_N/dT (simulation)
$\frac{df_N}{dT}$
0,02
0,01
0
C_V
C_V (simulation)
600
500
400
300
200
100
0
0
100
200
300
400
500
600
700
T (K)

FIGURE 5.

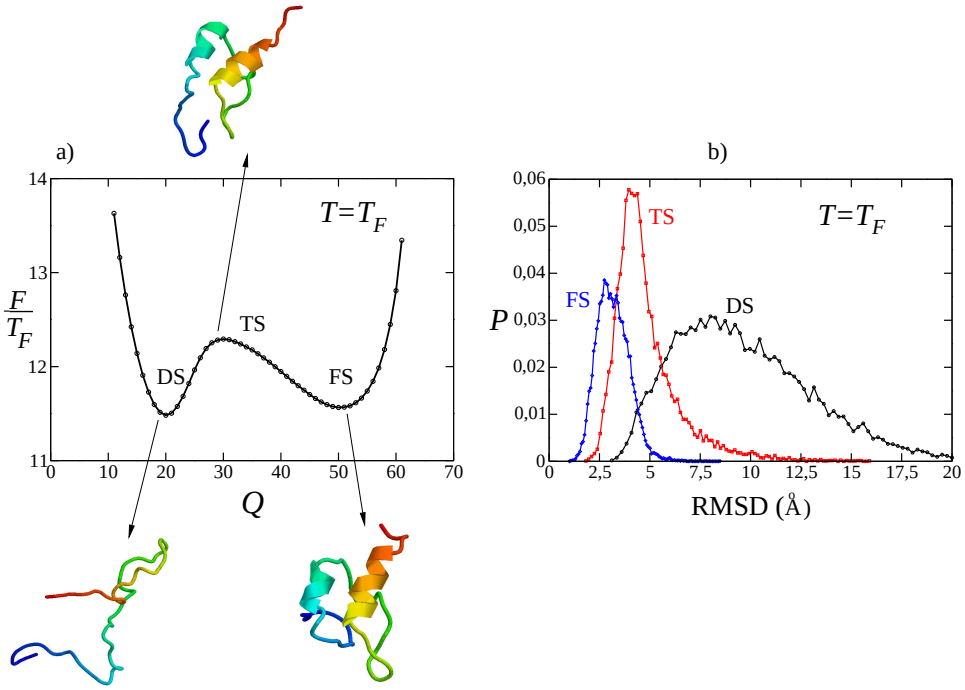

FIGURE 6.

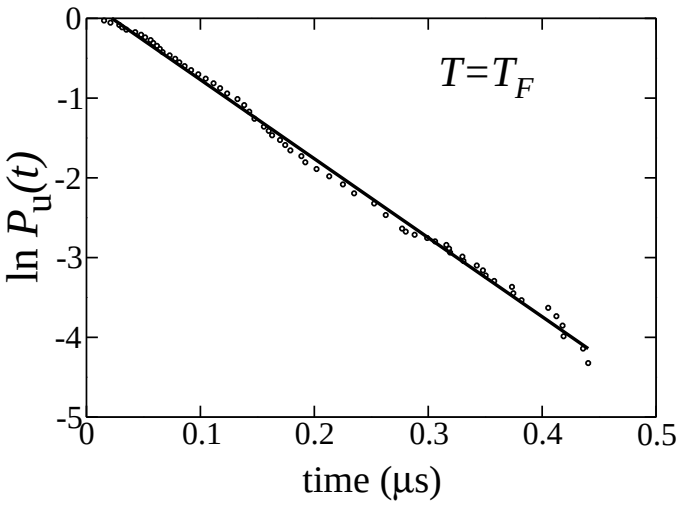

FIGURE 7.